

# Hacia la clasificación de actividad e inactividad lingüística a partir de señales de electroencefalogramas (EEG)

Luis Alfredo Moctezuma<sup>1</sup>, Maya Carrillo<sup>1</sup>, Luis Villaseñor Pineda<sup>2</sup>,  
Alejandro A. Torres García<sup>2</sup>

<sup>1</sup>Benemérita Universidad Autónoma de Puebla,  
Puebla, México

<sup>2</sup>Instituto Nacional de Astrofísica Óptica y Electrónica,  
Tonantzintla, Puebla, México

luisalfredomoctezuma@gmail.com, cmaya@cs.buap.mx, {villasen,  
alejandro.torres}@ccc.inaoep.mx

**Resumen.** En este artículo se evalúan dos formas de caracterización de señales de EEG para clasificación de actividad e inactividad lingüística, preproceso indispensable para el desarrollo de interfaces cerebro-computadora (BCI) basadas en habla imaginada. La primera caracterización utilizó el cálculo de la energía *teager* con base en la transformada discreta wavelet (DWT). La segunda caracterización se basó en el cálculo de valores estadísticas directamente sobre la señal. Los experimentos se realizaron sobre dos bases de datos de habla imaginada, una con 27 sujetos y otra con 20. Los resultados muestran que al usar los valores estadísticos, la actividad e inactividad lingüística se distingue mejor que al usar la DWT; sin embargo, cabe señalar que con ambas caracterizaciones se alcanzan altas tasas de exactitud.

**Palabras clave:** habla imaginada, interfaces cerebro computadora, electroencefalogramas (EEG), transformada discreta wavelet (DWT), características estadísticas.

## Towards Classification of Activity and Linguistic Inactivity from Electroencephalograms (EEG) Signals

**Abstract.** This article evaluates two ways of characterization of EEG signals for classification of activity and linguistic inactivity, essential preprocess for the development of brain-computer interfaces (BCI) based on imagined speech. The first characterization used the calculation of the *teager* energy based on the discrete wavelet transform (DWT). The second characterization was based on the calculation of statistical

values directly from the signal. The experiments were performed on two imagined speech databases, the first one with 27 subjects and the other one with 20. The results show that when using statistical values, linguistic inactivity and activity are better distinguished than when using DWT; however, it should be noted that with both characterizations high accuracy rates are reached.

**Keywords:** imagined speech, brain computer interfaces, electroencephalograms (EEG), discrete wavelet transform (DWT), statistical characteristics.

## 1. Introducción

La organización mundial de la salud (OMS) define el término de discapacidad como todas las deficiencias, limitaciones y las restricciones de la participación en situaciones vitales, por ejemplo, la comunicación hablada. Datos del Instituto Nacional de Estadística y Geografía<sup>1</sup> (INEGI) del 2015 muestran que el 6 % de la población mexicana sufre alguna discapacidad, de las cuales el 18 % sufre alguna discapacidad para hablar o comunicarse; entre las más destacadas se encuentran la esclerosis lateral amiotrófica, esclerosis múltiple, lesiones de médula espinal o cerebral, entre otras. Las personas que sufren dichas discapacidades tienen muy diversos retos al convivir día a día con las personas que los rodean, no pueden desempeñar algunas tareas y no pueden comunicar lo que piensan o sienten. Una solución a esta situación es el uso de interfaces cerebro computadora (BCI, por sus siglas en inglés) que brindan un nuevo canal de comunicación a personas con este tipo de discapacidades.

Una BCI es un sistema de comunicación que monitorea la actividad cerebral y traduce determinadas características, correspondientes a las intenciones del usuario, en comandos de control de un dispositivo. Las BCI se pueden dividir en invasivas y no invasivas. Las primeras requieren de procedimientos quirúrgicos y se obtienen señales cerebrales muy nítidas gracias a que las medidas no son atenuadas por el cráneo y el cuero cabelludo. Las segundas, las BCI no invasivas, no requieren ningún tipo de cirugía, sin embargo, las señales que se obtiene son más débiles. Dentro de las no invasivas las basadas en electroencefalogramas son las más utilizadas debido a su costo relativamente bajo.

Las fuentes electrofisiológicas se refieren a los mecanismos o procesos neurológicos usadas por un usuario de una BCI para generar señales de control[2]. Wolpaw[3] separó estas fuentes en 5 categorías basadas en los mecanismos neuronales y la tecnología de grabación que utilizan. Estas categorías son: actividad sensoriomotora, potenciales P300, potenciales evocados visuales (VEP), potenciales corticales lentos (SCP) y actividad de la célula neural (ANC). Luego Bashashati[4] agregó 2 categorías más: respuesta a tareas mentales y múltiples neuromecanismos.

<sup>1</sup> Estadísticas a propósito del día internacional de las personas con discapacidad realizado por el INEGI en México (2015)

Sin embargo, dichas fuentes electrofisiológicas presentan dos problemas: (i) el largo periodo de entrenamiento, pues las fuentes electrofisiológicas son generadas por el usuario de manera poco consiente; y (ii) las bajas tasas de comunicación, una sola palabra o menos procesada por minuto, lo que hace imposible la comunicación de manera natural. Es por esto que Desain[5] dice que aunque prometedor, todavía se está lejos de lo que se necesita para un control rápido y fiable de las interfaces y que en lugar de tratar de mejorar los métodos que usan las fuentes electrofisiológicas anteriores, se deben aprovechar otras tareas mentales. Entre ellas menciona el habla imaginada o interna; que se refiere a la pronunciación interna, o imaginada, de palabras pero sin emitir sonidos ni articular gestos.

El habla imaginada como fuente electrofisiológica tiene ventajas respecto a las otras, pues necesita poco entrenamiento; sin embargo, es una tarea compleja, pues está ligada a múltiples artefactos tales como movimiento de músculos y pulsos de arterias en el cuero cabelludo. En este trabajo se explora lo que sería un primer paso de una BCI utilizando EEG para habla imaginada : distinguir la actividad cerebral generada por el sujeto al imaginar una palabra (i.e. emprende una actividad lingüística) de cualquier otra actividad cerebral. Los experimentos realizados buscan evaluar diferentes formas de representación de la señal para realizar dicha distinción. Las señales de EEG fueron separadas en 2 clases, un conjunto de palabras imaginadas consideradas la clase de *actividad lingüística*; y los estados de reposo o pausa como ejemplos de otra actividad cerebral. Estos últimos denominados por simplicidad *inactividad lingüística*.

## 2. Trabajos relacionados

A continuación se presenta una revisión de los trabajos realizados en la tarea de reconocimiento de habla no pronunciada o imaginada. En algunos de ellos se han reportado experimentos enfocados a reconocer sólo sílabas o fonemas para el control de algún dispositivo, y otros consideran un enfoque orientado a reconocer palabras imaginadas. Cabe señalar que no todos estos trabajos intentan detectar el inicio del habla imaginada.

Respecto al reconocimiento de palabras imaginadas, en [11] se grabaron 20 repeticiones de las palabras imaginadas: *alpha, bravo, charlie, delta, echo* de 18 sujetos. Para extraer sus características se utilizó *Double-Tree Complex Wavelet Transform (DTCWT)*, y para la etapa de clasificación *offline* se utilizó Janus recognition toolkit(Jrtk). La tasa de reconocimiento promedio alcanzada fue de 45.50%. En la discusión presentada por los autores, se destacó que ciertos inconvenientes afectan en general a los métodos de caracterización y clasificación, el primero es que cuanto mayor sea el tiempo de reposo, la tasa de reconocimiento disminuirá; y el segundo es que sólo los bloques de palabras imaginadas que fueron grabadas en la misma sesión obtendrán tasas de reconocimiento significativas. Estos puntos deben tenerse en cuenta para crear un dispositivo que pueda reconocer palabras imaginadas en tiempo real.

En [9] se hizo una aproximación hacia detectar el inicio del habla imaginada vs. estados de inactividad, en un estado de relajación y con el menor ruido posible en el área. En dicho trabajo se realizaron 3 tareas con 4 personas (3 hombres y una mujer) para obtener información lingüística. La primera tarea, llamada *overt*, consiste en obtener la señal cerebral cuando el sujeto imaginaba la palabra. La segunda tarea, llamada *I\_overt*, se exhalaba y tensaban las cuerdas vocales sin pronunciar el sonido; y la tercera tarea, llamada *covert*, consistió en imaginar y escuchar el sonido. La única palabra que imaginaron los sujetos fue la palabra *um* y se hicieron 6 registros con cada uno de los participantes.

Cada registro es una secuencia aleatoria de entre las tres tareas, por ejemplo: *over-covert-covert-I\_overt...* etc. Cada tarea fue de 5 segundos, y después de procesar los datos se tomaron ventanas de tamaño 1,1.5, 2 y 3 segundos para extraer características usando un modelo de regresión lineal (Método de Burg). Lo anterior fue realizado con el fin de comprobar si las tareas planteadas se encontraban en una ventana menor a los 5 segundos de cada instancia. Para la clasificación se usó análisis de discriminación lineal que busca una línea que maximiza la separación entre clases y minimiza el traslape entre ellas. Los resultados muestran que la tarea 3 (*covert*) es la que mejor se pudo discriminar alcanzando para cada individuo 82.4 % 81.20 % 85.12 % y 74.72 % de precisión respectivamente.

En [10] se utilizaron las vocales *a* y *u* para mostrar la factibilidad de controlar una prótesis utilizando la imaginación de dichas vocales, y se obtuvieron resultados de 68 % y 78 % de exactitud para cada vocal. Este trabajo consistió de un framework que realiza el procedimiento de clasificación de las vocales en cuestión de segundos. Cabe mencionar, que se aprovechó la activación de las vocales seleccionadas *a* y *u* con diferentes músculos. Para este trabajo se usó Biosemi ActiView capturando 64 canales EEG con una frecuencia de muestreo de 2048 Hz. Se hicieron ventanas de 500 ms y se obtuvieron características con patrones espaciales comunes (*common spatial pattern* CSP) para clasificar con máquinas de soporte vectorial con una función de base radial.

En [8] se utilizó la DWT para extraer características y éstas usarlas como entrada para 4 clasificadores: Naive Bayes, Máquinas de soporte vectorial, Random Forest, y Perceptrón multicapa para una base de datos de 120 ejemplos de las palabras *arriba* y *abajo* obtenidas de 3 personas. Se usó el kit emotiv EPOC para obtener los datos, en este trabajo se estudiaron más a profundidad los canales cercanos al área de Wernicke, que corresponden a los canales F7, FC5, T7 y P7 en el sistema internacional 10-20.

Se discute que la clasificación de las palabras imaginadas es independiente del clasificador ya que no se muestran cambios significativos entre los clasificadores usados, sin embargo, los mejores resultados son con Random Forest. Además se da una aproximación de cómo se podrían mejorar los resultados, aplicando análisis de componentes independientes y filtrando los componentes mediante el exponente de Hurst.

Los trabajos antes mencionados hacen uso de habla imaginada, sin embargo, en [7] se incluyen expresiones faciales y audio. En dicho trabajo

participaron 4 mujeres y ocho hombres que pronunciaron e imaginaron las sílabas /iy//uw//piy//tiy//diy//m//n//pat//pot//knew//gnaw/. Se realizaron 132 ensayos, después de cada 40 ensayos los participantes descansaban. En esta investigación se calcularon medidas estadísticas para cada ventana: media, mediana, desviación estándar, varianza, máximo, mínimo, entropía espectral, energía, curtosis y oblicuidad. Para la clasificación emplearon máquinas de soporte vectorial y se combinó información tanto de EEG, como expresiones faciales y audio.

En este trabajo presentamos 2 métodos para la caracterización de la señal de EEG y con ello distinguir entre actividad e inactividad lingüística, sin importar la palabra imaginada. De esta forma, una vez identificado un segmento de actividad lingüística un proceso posterior identificará la palabra imaginada.

### 3. Método

Nuestro método consiste en 3 etapas: preprocesar los datos, extraer las características, y crear el clasificador. Dichas etapas se detallan en las secciones a continuación.

Cabe señalar que para realizar la clasificación se utilizó Random Forest pues ha mostrado lo mejores resultados [8]. Las características fueron calculadas usando la DWT bior 2.2 con diferentes niveles de descomposición y a cada uno de los niveles se le calcularon varios tipos de energía con el fin de compararlos. También se extrajeron otro tipo de características, basadas en información estadística sobre la señal cruda, retomando algunas de las presentadas en [7].

En general, el procedimiento seguido se muestra en la figura 1.



Fig. 1. Procedimiento general.

#### 3.1. Pre-procesamiento

Esta etapa se refiere a la reducción y transformación de las bases de datos con el fin de tener datos adecuados con los cuales trabajar. Para pre-procesar las señales EEG se usa el método de referencia promedio común (CAR). Este método mejora la relación señal-ruido de la señal de EEG. El CAR se calcula

mediante la siguiente formula:

$$V_i^{CAR} = V_i^{ER} - \frac{1}{n} \sum_{j=1}^n V_j^{ER}, \quad (1)$$

donde  $V_i^{ER}$  es el potencial entre el  $i$ -ésimo electrodo y la referencia, y  $n$  es el número de electrodos.

### 3.2. Extracción de características

Para obtener características se calcularon coeficientes de energía con DWT o medidas estadísticas. En ambos casos fueron calculadas las características para cada uno de los canales y posteriormente concatenados, formando así un sólo vector de características para cada instancia o época.

**DWT** Las señales de EEG generalmente son no estacionarias ya que cambian rápidamente con el tiempo. Además, estas características contienen información del tiempo debido a que los patrones de actividad cerebral están generalmente relacionados a variaciones específicas del EEG en el tiempo. Por esa razón se debe usar una representación que considere este comportamiento.

Como primera opción se aplicó la DWT con la función madre *bior 2.2* a cada uno de los canales, con diferentes niveles de descomposición. Al aplicar dicha transformada con un nivel de descomposición  $j$  se obtiene una estructura con vectores de coeficientes de aproximación ( $CA_j$ ) y detalle ( $CD_j$ ):  $[CA_j, CD_j, CD_{j-1}, \dots, CD_1]$ . Sin embargo, el número de coeficientes wavelet en cada nivel de descomposición variará dependiendo la duración de la ventana de la pronunciación de la palabra imaginada (tanto entre palabras de un mismo sujeto como entre palabras de distintos sujetos). Para tratar dicho problema se calculó la distribución de varios tipos de energía (Instantánea, Jerárquica, Relativa y Teager)[4] en cada nivel, de esta manera se logra normalizar todas las instancias de EEG. A continuación se muestran las formulas de los diferentes tipos de energía.

Sea  $w_j(r)$  el coeficiente wavelet en el tiempo  $r$  y banda de frecuencia  $j$ . Si  $N$  es la longitud de la ventana de análisis,  $w_j(r)$  tiene  $N_j = N/2^j$  muestras en el nivel de descomposición  $j$ :

- Energía Instantánea: este tipo de energía refleja la amplitud de la señal.

$$f_j = \log_{10} \left( \frac{1}{N_j} \sum_{r=1}^{N_j} (w_j(r))^2 \right). \quad (2)$$

- Energía Jerárquica: provee una resolución jerárquica en tiempo y le da más importancia al centro de la ventana de análisis.

$$f_j = \log_{10} \left( \frac{1}{N_j} \sum_{r=(N_j-N_j)/2}^{(N_j+N_j)/2} (w_j(r))^2 \right), \quad (3)$$

donde además  $N_j$  es el número de muestras en el último nivel de descomposición.

- Energía Teager: refleja las variaciones tanto en amplitud como en frecuencia de la señal y es un parámetro robusto para el reconocimiento de habla ya que atenúa el ruido auditivo.

$$f_j = \log_{10} \left( \frac{1}{N_j} \sum_{r=1}^{N_j-1} |(w_j(r))^2 - w_j(r-1) * w_j(r+1)| \right). \quad (4)$$

- Energía Relativa: la distribución de energía puede ser considerada como una densidad tiempo-escala y provee información para caracterizar la distribución de energía de la señal en diferentes bandas de frecuencia.

$$f_j = \frac{E_j}{E_{total}}, \quad (5)$$

donde  $E_j$  representa la energía en el  $j$ -ésimo nivel de descomposición y  $E_{total}$  representa la energía total de los coeficientes wavelet de una señal dada. La energía en el  $j$ -ésimo nivel de descomposición  $E_j$  se define como:

$$E_j = \begin{cases} \sum_k |d_{j,k}|^2 & \text{Si } j \leq N, \\ \sum_k |a_k|^2 & \text{en caso contrario.} \end{cases} \quad (6)$$

Y la energía total  $E_{total}$  como sigue:

$$E_{total} = \sum_{j=1}^{N+1} E_j. \quad (7)$$

Estos coeficientes de energía fueron calculados para cada electrodo y concatenados para formar un vector de características que representan a la señal de EEG.

**Estadísticas** Otra opción para la caracterización de la señal fue motivada por lo planteado en [7]. En dicho trabajo se caracterizó la señal al calcular un conjunto de valores estadísticos. Inicialmente tomamos 15 características estadísticas por canal. Es decir, para cada uno de los canales se calcularon las 15 características y luego se concatenaron, para tener un vector de características por cada instancia. Como se verá más adelante se usaron dos conjuntos de valores estadísticos.

### 3.3. Creación del clasificador

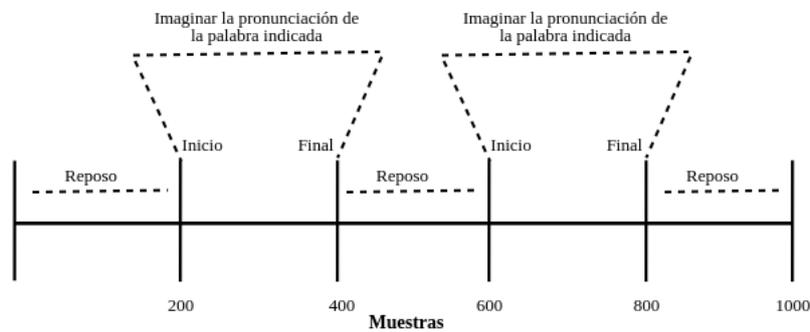
Una vez obtenidos los vectores de características para cada muestra de EEG se utilizó un algoritmo de clasificación automática para calcular un modelo que distinga entre actividad e inactividad lingüística. Para ello se utilizó el algoritmo de *Random Forest*. Este algoritmo fue seleccionado debido a los resultados reportados en trabajos relacionados. Para evaluar el rendimiento del clasificador se usó la exactitud.

## 4. Experimentos y resultados

El objetivo de esta sección es mostrar los resultados al experimentar con los dos conjuntos de características diferentes para clasificar en actividad e inactividad. Cabe recordar que la clase de actividad lingüística está conformada por las instancias de las palabras imaginadas y la inactividad esta formada por pausas o reposos. A continuación se detallan las bases de datos usadas, para posteriormente mostrar los resultados alcanzados al utilizar diferentes tipos de caracterizaciones.

### 4.1. Datos

Para los experimentos se utilizaron dos bases de datos. La primera base de datos (BD1) contiene las señales de EEG de 27 sujetos hablantes nativos del idioma español registradas usando EMOTIV EPOC mientras imaginaban 33 repeticiones de 5 palabras (*arriba*, *abajo*, *izquierda*, *derecha* y *seleccionar*) en idioma español, separadas por un estado de reposo, tal como se muestra en la figura 2 y se describe en [1]. Para cada palabra imaginada se obtuvo y guardo información de 14 canales de EEG con una frecuencia de muestreo de 128 Hz organizados de acuerdo al sistema internacional 10-20.

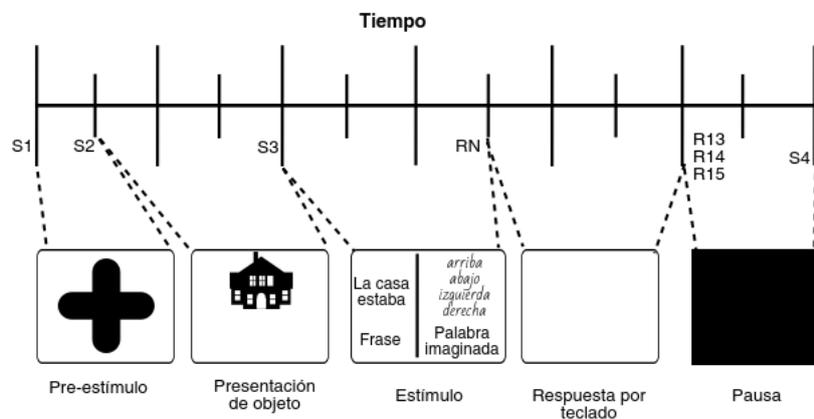


**Fig. 2.** Diseño del protocolo para adquisición de la señal de EEG para la BD1 usando EMOTIV EPOC.

Para la BD1 la clase actividad contiene 165 instancias que corresponden a las 33 repeticiones de cada una de las 5 palabras. La clase inactividad contiene 165 instancias que corresponden a los estados de reposo que existen entre una repetición de una palabra y otra. Eso para cada uno de los 27 sujetos.

La segunda base de datos (BD2) contiene las señales de EEG de 20 sujetos hablantes nativos del idioma español que fueron grabados mientras imaginaban 40 repeticiones de 4 palabras (*arriba*, *abajo*, *derecha*, *izquierda*) separadas por una pausa, así como se describe en [1].

El protocolo de estimulación y adquisición de las señales de EEG de la BD2 se muestra en la figura 3, donde  $RN$  es el marcador final para  $N = 1, 2, 3, 4$  correspondientes a las 4 palabras imaginadas:  $R1$  para la palabra imaginada *arriba*,  $R2$  para *abajo*,  $R3$  para *derecha* y  $R4$  para *izquierda*. El marcador  $\{R13$  o  $R14$  o  $R15\}$  es el final de la respuesta por teclado que indica el momento en que el sujeto presionaba una flecha en el teclado y dicha flecha debería coincidir con la palabra que se solicitó imaginar. El tiempo máximo de espera para imaginar la palabra fue de  $1,5s$ , a menos que el sujeto presionara la tecla antes. Además  $\{R13$  o  $R14$  o  $R15\}$  es el inicio de una pausa terminando en el marcador  $S4$ .



**Fig. 3.** Diseño del protocolo para estimulación y adquisición de la señal de EEG para la BD2 usando Brain Vision.

Para cada palabra imaginada se guardó información de 64 canales de EEG con una frecuencia de muestreo de 500 Hz. Para la BD2 la clase actividad contiene 160 instancias que corresponden a las 40 repeticiones de cada una de las 4 palabras imaginadas (la frase que precede a la palabra imaginada también se consideró como actividad) entre los marcadores  $S3$  a  $RN$  según la figura 3 y la clase inactividad contiene 160 muestras que corresponden a 40 instancias de el pre-estímulo, 40 instancias de la presentación del objeto, 40 instancias de la respuesta por teclado y finalmente 40 instancias de la pausa (ver figura 3).

La tabla 1 presenta una descripción de las principales características de las bases de datos utilizadas.

#### 4.2. Características usando DWT

Las características calculadas con DWT bior2.2 fueron realizadas con distintos niveles de descomposición. Para describir la descomposición se parte de conocer la frecuencia de muestreo ( $fN$ ) que es de 128 Hz para la BD1 y 500 Hz

**Tabla 1.** Comparación de las características de las bases de datos usadas.

	BD1	BD2
Num. sujetos	27	20
Num. palabras imaginadas	5	4
Num. repeticiones por palabra	33	40
Canales de EEG	14	64
Frecuencia de muestreo	128 Hz	500 Hz

para la BD2. Luego, el teorema de Nyquist dice que la frecuencia de muestreo debe ser por lo menos 2 veces mayor a la frecuencia de la señal a muestrear, es decir:  $f_N \geq 2f_s$ , donde  $f_s$  es la frecuencia de la señal muestreada. Despejando  $f_s$ , se obtiene:  $f_s \approx \frac{f_N}{2}$ . Por lo tanto para la BD1 se tiene que  $f_s \approx 64Hz$ . y para la BD2  $f_s \approx 250Hz$ .

Por otro lado, también es necesario observar el tamaño de las muestras para determinar el máximo número de niveles de descomposición que pueden ser calculados. La tabla 2 muestra los posibles niveles de descomposición para cada base de datos.

**Tabla 2.** Rangos de frecuencias(Hz) para el nivel de descomposición 4 de DWT para la BD1 y los niveles de descomposición 5 y 6 para la BD2.

BD1		BD2			
Nivel	Rango	Nivel	Rango	Nivel	Rango
D1	32-64	D1	125-250	D1	125-250
D2	16-32	D2	62.5-125	D2	62.5-125
D3	8-16	D3	31.25-62.5	D3	31.25-62.5
D4	4-8	D4	15.62-31.25	D4	15.62-31.25
A4	0-4	D5	7.81-15.62	D5	7.81-15.62
		A5	0-7.81	D6	3.9-7.81
				A6	0-3.9

Para seleccionar el conjunto de coeficientes DWT más adecuado para la señal de EEG se consideró lo planteado en [6,11] donde menciona que las frecuencias de EEG llegan hasta 60 Hz y más allá de dichas frecuencias no tienen interés para nuestro análisis. Tomando en cuenta lo anterior se realizó la caracterización para los 27 sujetos de la BD1 y los 20 sujetos de la BD2 con la DWT bior 2.2 con 4 niveles de descomposición para la BD1 y con 5 y 6 para la BD2 para luego eliminar los coeficientes de detalle correspondientes a los rangos mayores a 60 Hz.

Los resultados promedio se muestran en la tabla 3. Para la BD1 sólo se pueden calcular 4 niveles de descomposición pues la frecuencia de muestreo es de 128 Hz. Para la BD2 la frecuencia de muestreo es de 500 Hz, por lo tanto

**Tabla 3.** Exactitud y desviación estándar del clasificador Random Forest a 10 pliegues para DWT bior2.2 con 4 niveles de descomposición usando la BD1 al calcular diferentes tipos de energía.

Nivel	Num. Características	Instantánea	Jerárquica	Relativa	Teager
4	70, 5 por canal	0.75 ±0.15	0.73 ±0.14	0.73 ±0.16	0.78 ±0.16

se puede tener un mayor número de niveles de descomposición. Si se calculan niveles de descomposición menores a 5 sólo se obtienen 2 características por canal después de eliminar los rangos más allá de 60 Hz, es por eso que se deben calcular las características con niveles de descomposición mayores a 5. Además se debe considerar que el máximo nivel de descomposición de esta base de datos es 6. A continuación se muestran los resultados promedio de exactitud para los niveles de descomposición 5 y 6 para la BD2.

**Tabla 4.** Exactitud y desviación estándar del clasificador Random Forest a 10 pliegues para DWT bior2.2 con 5 y 6 niveles de descomposición usando la BD2 al calcular diferentes tipos de energía. El sub-índice en el Nivel indica los niveles de detalle eliminados.

Nivel	Num. Características	Instantánea	Jerárquica	Relativa	Teager
5 <sub>2</sub>	256, 4 por canal	0.90 ±0.02	0.90 ±0.02	0.89 ±0.03	0.91 ±0.03
6 <sub>2</sub>	320, 5 por canal	0.90 ±0.02	0.90 ±0.01	0.89 ±0.03	0.92 ±0.03

En la tabla 4, los sub-índices  $n = 2$  para los niveles de descomposición indican que se han eliminado los primeros  $n$  coeficientes de detalle para evitar las frecuencias más allá de 60 Hz. Como se puede observar en la tabla 3 y 4 la caracterización usando la energía Teager es la que presenta mejores resultados.

### 4.3. Características estadísticas

Las características utilizadas para este experimento constan de 2 conjuntos: el primero es de 15 valores estadísticos y el segundo de 9. El primer conjunto(C1) son las características usadas por [7] que consta de 9 valores estadísticos y 6 combinaciones de algunos de ellos. En el segundo(C2) conjunto sólo usamos los 9 valores sin considerar las combinaciones, como se muestra a continuación.

- C1: media, media<sup>6</sup>, absoluto de la media, máximo, absoluto del máximo, mínimo, absoluto del mínimo, máximo+mínimo, máximo-mínimo, desviación estándar, varianza, curtoris, oblicuidad, suma y mediana.
- C2: media, máximo, mínimo, desviación estándar, varianza, curtoris, oblicuidad, suma y mediana.

En la tabla 5 se muestran los resultados de las 2 base de datos al usar los valores estadísticos como características. En los 2 casos se obtienen los mejores resultados al usar 9 características por canal.

**Tabla 5.** Exactitud y desviación estándar del clasificador random forest a 10 pliegues para características estadísticas usando las 2 bases de datos.

BD	Características	Num. Características	Exactitud
BD1	C1	210, 15 por canal	0.82 ±0.18
BD1	C2	126, 9 por canal	0.83±0.17
BD2	C1	960, 15 por canal	0.91 ±0.05
BD2	C2	576, 9 por canal	0.91±0.05

Como puede observarse en los resultados reportados, la caracterización usando el conjunto de 9 características(C2) son mejores o iguales que los obtenidos al usar la DWT. En la tabla 6 se muestra el comparativo de los resultados alcanzados para los 27 sujetos de la BD1 con los dos tipos de caracterización (utilizando DWT y características estadísticas). Utilizando DWT con la función madre bior 2.2, los mejores resultados se obtuvieron con 4 niveles de descomposición y calculando la energía Teager a cada uno de los niveles. Para las características estadísticas los mejores resultados se obtuvieron al usar el conjunto de 9 características(C2) por canal.

**Tabla 6.** Exactitud, media y desviación estándar (Std) del clasificador random forest a 10 pliegues para características DWT calculando la energía Teager y estadísticas(EST) para la BD1.

Sujeto	DWT	EST	Sujeto	DWT	EST	Sujeto	DWT	EST
1	0.76	0.83	10	0.98	1.00	19	0.99	1.00
2	0.90	0.99	11	0.64	0.74	20	0.81	0.93
3	0.67	0.66	12	0.77	0.84	21	0.88	0.94
4	0.74	0.79	13	0.46	0.57	22	0.93	0.98
5	0.97	1.00	14	0.57	0.63	23	0.57	0.53
6	0.90	0.95	15	0.90	0.96	24	0.86	0.92
7	0.92	0.91	16	0.75	0.76	25	0.98	1.00
8	0.90	0.93	17	0.52	0.52	26	0.77	0.87
9	0.66	0.77	18	0.48	0.42	27	0.85	0.97
<b>Media</b>							<b>0.78</b>	<b>0.83</b>
<b>Std</b>							<b>± 0.16</b>	<b>± 0.17</b>

En la tabla 7 se muestran los mejores resultados para los 20 sujetos de la BD2 tanto usando DWT como características estadísticas. Para el caso de DWT se usaron 6 niveles de descomposición, se eliminaron los primeros 2 coeficientes

de detalle y se calculó la energía Teager para los restantes 5 coeficientes (D3, D4, D5, D6 y A6). Esto se repitió por cada canal. Con las características estadísticas se utilizó el conjunto de características C2.

**Tabla 7.** Exactitud, media y desviación estándar (Std) del clasificador Random Forest a 10 pliegues para características DWT calculando la energía Teager y estadísticas(EST) para la BD2.

Sujeto	DWT	EST	Sujeto	DWT	EST	Sujeto	DWT	EST
1	0.99	1.00	8	0.95	0.98	15	0.91	0.88
2	0.87	0.86	9	0.92	0.92	16	0.89	0.88
3	0.88	0.88	10	0.92	0.92	17	0.94	0.94
4	0.90	0.84	11	0.93	0.93	18	0.94	0.96
5	0.88	0.88	12	0.89	0.89	19	0.93	0.96
6	0.89	0.84	13	0.92	0.93	20	0.96	0.94
7	0.88	0.88	14	0.94	0.96			
<b>Media</b>							<b>0.92</b>	<b>0.91</b>
<b>Std</b>							<b>± 0.03</b>	<b>± 0.05</b>

## 5. Conclusiones y trabajo futuro

El reconocimiento de habla imaginada, a partir de señales de EEG, ha sido estudiado en sistemas fuera de línea como una tarea de clasificación de distintas palabras de un vocabulario propuesto. En este trabajo se presentó un estudio para identificar la actividad e inactividad lingüística con el fin de dar el primer paso hacia el reconocimiento en línea. Los experimentos realizados a pesar de también llevarse a cabo fuera de línea dan evidencia de la factibilidad de la tarea. Se analizó si, independientemente de la palabra imaginada (actividad lingüística) es posible diferenciarse de otras actividades cerebrales (inactividad lingüística).

El método propuesto fue evaluado con dos bases de datos, con diferente protocolo de registro y diferentes instrumentos. Los resultados son muy alentadores, indicando que es posible distinguir entre actividad o inactividad lingüística. Los resultados alcanzados están muy por arriba de la cota del azar, lo que permite conjeturar que es posible crear un proceso en línea que logre clasificar una señal como actividad o inactividad y posteriormente con otro clasificador identificar la palabra imaginada que finalmente se convertirá en un comando para fines de comunicación o control de dispositivos.

Por otra parte, este trabajo evaluó si el tipo de caracterización de las señales de EEG tiene algún efecto en la discriminación de actividad e inactividad lingüística. Las caracterizaciones estudiadas fueron características estadísticas obtenidas desde las señales en tiempo y energías calculadas a partir de la DWT. Usando los vectores estadísticos se obtuvieron resultados iguales o superiores

que con las energías wavelet; con la ventaja de que éstos no requieren de una transformación a otro dominio, ni definición de parámetros adicionales a la señal.

Al comparar los resultados de las tablas 6 y 7 se puede notar que siguiendo el mismo procedimiento en ambas bases de datos se obtiene mayor exactitud en la BD2. Son varios factores por los cuales la exactitud es mayor para la BD2. Como puede verse en la tabla 1 la BD2 comprende un número mayor de repeticiones de cada palabra, una frecuencia de muestreo mayor y se cuenta con información de un mayor número de canales (64 canales).

Para comprobar los resultados obtenidos aquí, como trabajo futuro se planea probar otras bases de datos con protocolos de adquisición y frecuencia de muestreo diferentes. Además, se plantea dividir la señal de EEG y comprobar si aún existe información suficiente para distinguir entre actividad e inactividad lingüística. De esta manera se podrá identificar en un menor tiempo el inicio y fin de la actividad lingüística. Es así que será posible orientar los trabajos de reconocimiento de habla imaginada a tiempo real.

**Agradecimientos.** Los autores agradecen al Consejo Nacional de Ciencia y Tecnología (CONACyT) por al apoyo a esta investigación, el cual se realizó a través de la beca No. 591475 y del proyecto No. Ref. 2016-01-2228. Los autores también desean agradecer el apoyo de la Red Temática en Tecnologías del Lenguaje/CONACYT, No. Ref. 260178, 271622, por los mecanismos de colaboración brindados para realizar este trabajo.

## Referencias

1. Torres-García, A. A.: Análisis y clasificación de electroencefalogramas (EEG) registrados durante el habla imaginada. Tesis de doctorado, Instituto Nacional de Astrofísica, Óptica y Electrónica, México (2016)
2. Bashashati, A., Fatourehchi, M., Ward, R. K., Birch, G. E.: A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural Engineering*, vol. 4 (2007)
3. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T. M.: Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, vol. 113, pp. 767–791 (2002)
4. Didiot, E., Illina, I., Fohr, D., Mella, O.: A Wavelet-Based Parameterization for Speech/Music Discrimination. vol. 24, pp. 341, Vandoeuvre-les-Nancy, France (2009)
5. Desain, P., Farquhar, J., Haselager, P., Hesse, C., Schaefer, C.: What BCI research needs. In: *Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems*, vol. 55, pp. 181–184, Florence, Italy (2008)
6. Brigham, K., Vijaya Kumar, B.V.K.: Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy. In: *Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference on*, IEEE, pp. 1–4. Pittsburgh PA, USA (2010)
7. Zhao, S., Rudzicz, F.: Classifying Phonological Categories in Imagined and Articulated Speech. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 22, pp. 992–996, Toronto, Canada (2015)

8. Torres-García, A. A., Reyes-García, C. A., Villaseñor-Pineda, L.: Análisis de Señales Electroencefalográficas para la Clasificación de Habla Imaginada. *Revista Mexicana de Ingeniería Biomédica*, vol. 34, pp. 23–29 (2013)
9. Song, Y., Sepulveda, F.: Classifying speech related vs. idle state towards onset detection in brain-computer interfaces overt, inhibited overt, and covert speech sound production vs. idle state. In: *IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings*, vol. 55, pp. 568–571, Essex, United Kingdom (2014)
10. Kambara, H., Sato, M., DaSalla, C. S., Koike, Y.: Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Networks*, vol. 22, pp. 1334–1339 (2009)
11. Porbadnigk, A., Schultz, T.: Eegbased speech recognition: Impact of experimental design on performance. Master's thesis, Institut für Theoretische Informatik Universität Karlsruhe, Germany (2008)